4 | ESTUDO DE CASO

4.1 Seleção

Conforme a explicação anterior, a etapa de seleção envolve a compreensão do domínio e dos objetivos da tarefa a ser desenvolvida, bem como a obtenção dos dados (atributos/características). Para esse estudo de caso, os dados foram coletados a partir das respostas de um formulário disponibilizado para alunos que estudavam na escola X. As questões deste formulário continham:

1. Sexo (Masculino, Feminino)
2. Localidade de residência (Rural, Urbana)
3. Utiliza transporte escolar (Sim, Não)
4. Participa do projeto social bolsa família (Sim, Não)
5. Idade (6-10, 11-15, 16-20, 21-25, 26-30)
6. Houve abandono da escola alguma vez (Sim, Não)
7. Qual motivo (os) que ocasionou (ram) o abandono: (falta de perspectiva profissional, casamento, bullying, escola não atrativa, gravidez, trabalho ou desinteresse).
8. A partir dos dados, o objetivo geral foi compreendido a partir do perfil dos alunos que evadiram da escola estadual X alguma vez e os fatores que levaram este abandono.

4.2 Pré-processamento

Foram eliminados dados de alunos que não evadiram, pois, o objetivo era justamente compreender os fatores que levaram os alunos a desistirem de continuar estudando em algum momento. No total, o conjunto de dados continha dados de 200 alunos.

4.3 Transformação

Nesta etapa, os dados foram formatados para que pudessem ser lidos pela ferramenta Weka. Assim, os dados foram exportados em formato CSV (Commaseparatedvalues), em que cada dado apresenta-se separado por vírgula. A Figura 3 apresenta uma amostra dos dados coletados (06 exemplos) e carregado na ferramenta Weka.

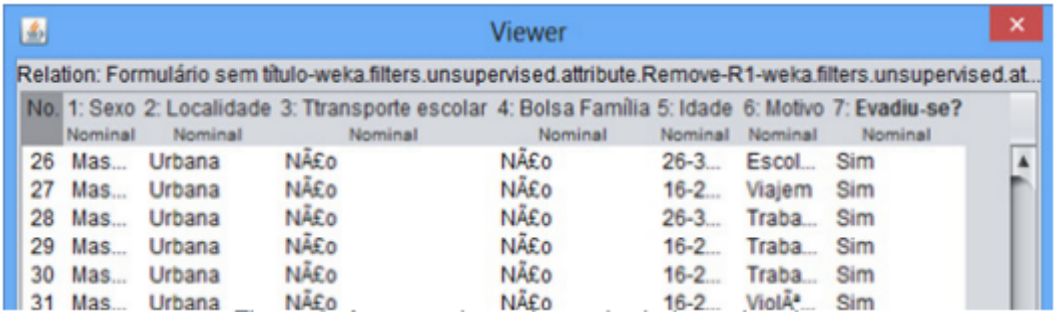

Figura 3. Amostra do conjunto de dados coletado.

Fonte: Autoria própria.

4.4 Mineração de dados

Nesta etapa foram utilizados os algoritmos de classificação Part, OneR, J48 e Randomtree, disponíveis na ferramenta Weka, para identificação de padrões (conhecimento). As cinco primeiras regras geradas por cada algoritmo são apresentadas nas Figuras 4, 5, 6 e 7.

R1: **SE** Idade 21-25 **ENTÃO** Motivo = Gravidez
R2: **SE** Idade 16-20 **ENTÃO** Motivo = Trabalho
R3: **SE** Idade 26-30 **ENTÃO** Motivo = Casamento
R4: **SE** Idade 6-10 **ENTÃO** Motivo = Trabalho
R5: **SE** Idade 11-15 **ENTÃO** Motivo = Trabalho

Figura 4. Regras geradas pelo algoritmo OneR

Fonte: autoria própria.

R1: **SE** Idade 26-30 **E** Bolsa Família = Não **E** Sexo = Masculino **ENTÃO** Motivo = Trabalho
R2: **SE** Idade 21-25 **ENTÃO** Motivo = Gravidez
R3: **SE** Idade 26-30 **ENTÃO** Motivo = Casamento
R4: **SE** Sexo = Masculino **ENTÃO** Motivo = Trabalho
R5: **SE** Bolsa Família = Não **ENTÃO** Motivo = Trabalho

Figura 5. Regras geradas pelo algoritmo Part

Fonte: autoria própria.

R1: **SE** Idade 21-25 **ENTÃO** Motivo = Gravidez
R2: **SE** Idade 16-20 **ENTÃO** Motivo = Trabalho
R3: **SE** Idade 26-30 **ENTÃO** Motivo = Casamento
R4: **SE** Idade 6-10 **ENTÃO** Motivo = Trabalho
R5: **SE** Idade 11-15 **ENTÃO** Motivo = Trabalho

Figura 6. Regras geradas pelo algoritmo J48

Fonte: autoria própria.

R1: **SE** Idade 21-25 **E** Bolsa Família = Sim **E** Residência = Urbana **ENTÃO** Motivo = Casamento
R2: **SE** Idade 21-25 **E** Bolsa Família = Sim **E** Residência = Rural **ENTÃO** Motivo = Trabalho
R3: **SE** Idade 21-25 **E** Bolsa Família = Não **E** Sexo = Masculino **ENTÃO** Motivo = Bullying
R4: **SE** Idade 21-25 **E** Bolsa Família = Não **E** Sexo = Feminino **E** Transporte Escolar = Não **ENTÃO** Motivo = Gravidez
R5: **SE** Idade 21-25 **E** Bolsa Família = Não **E** Sexo = Feminino **E** Transporte Escolar = Sim **ENTÃO** Motivo = Casamento

Figura 7. Regras geradas pelo algoritmo Randomtree

Fonte: autoria própria.

4.5 Avaliação

Esta etapa destinou-se a interpretação e avaliação dos resultados gerados na etapa anterior. Pôde-se verificar que as regras geradas pelo algoritmo OneR, conforme Figura 4, idade foi o único atributo utilizado para diferenciar o motivo da evasão. Na maioria dos casos, aponta trabalho como motivo da evasão, exceto para as idades de 21 a 30. Em relação as regras geradas pelo algoritmo Part, apresentadas na Figura 5, observou-se que alunos do sexo masculino e que não recebem bolsa família evadem tendo como motivo o trabalho. O algoritmo J48, por sua vez, também identificou trabalho como sendo o motivo principal para a evasão, exceto para as idades de 21-25 (gravidez) e 26-30 (casamento), conforme mostra a Figura 6. O algoritmo Randomtree gerou regras com o maior nível de detalhe, como pode ser visto na Figura 7. Percebeu-se que para a faixa etária de 21-25, os motivos podem ser variados (casamento, trabalho, bullying ou gravidez).

Revisão #1
Criado 7 outubro 2021 14:40:44 por Valerio Augusto Lopes Passos
Atualizado 7 outubro 2021 14:42:50 por Valerio Augusto Lopes Passos